CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Greenwich Academic Literature Archive

Greenwich Academic Literature Archive (GALA)
– the University of Greenwich open access repository
http://gala.gre.ac.uk

Citation for published version:

Moscati, Ivan and Tubaro, Paola (2011) Becker random behavior and the as-if defense of rational choice theory in demand analysis. Journal of Economic Methodology , 18 (2). pp. 107-128. ISSN 1350-178X (Print), 1469-9427 (Online)

Publisher's version available at:
http://dx.doi.org/10.1080/1350178X.2011.579147

Please note that where the full text version provided on GALA is not the final published version, the version made available will be the most up-to-date full-text (post-print) version as provided by the author(s). Where possible, or if citing, it is recommended that the publisher's (definitive) version be consulted to ensure any subsequent changes to the text are noted.

Citation for this version held on GALA:

Moscati, Ivan and Tubaro, Paola (2011) Becker random behavior and the as-if defense of rational choice theory in demand analysis. London: Greenwich Academic Literature Archive.
Available at: http://gala.gre.ac.uk/4660/

Contact: gala@gre.ac.uk

Ivan Moscati

Bocconi University
Department of Economics

Paola Tubaro

University of Greenwich
Business School

BECKER RANDOM BEHAVIOR AND THE AS-IF DEFENSE OF RATIONAL CHOICE THEORY IN DEMAND ANALYSIS

Forthcoming in the *Journal of Economic Methodology*

Abstract

In discussing rational choice theory (RCT) as an explanation of demand behavior, Becker (1962) proposed a model of random choice in which consumers pick up a bundle on their budget line according to a uniform distribution. This model has then been used in various ways to assess the validity of RCT and to support as-if arguments in defense of it.

The paper makes both historical and methodological contributions. Historically, it investigates how the interpretation of Becker random behavior evolved between the original 1962 article and the modern experimental literature on individual demand, and surveys six experiments in which it has been used as an alternative hypothesis to RCT.

Methodologically, the paper conducts an assessment of the as-if defense of RCT from the standpoint of Becker's model. It argues that this defense is "weak" in a number of senses, and that it has influenced negatively the design of experiments about RCT.

Keywords: rational choice; random choice; as-if defense of scientific theories; experimental economics; demand analysis.

JEL codes: B410, C140, D110, D120.

Correspondence may be addressed to Ivan Moscati at ivan.moscati@unibocconi.it or Paola Tubaro at p.tubaro@greenwich.ac.uk. We are grateful to Richard Bradley, Marco Dardi, John Davis, Franz Dietrich, Francesco Guala, Jean-Sébastien Lenfant, Marco Mariotti, Aldo Montesano, Jack Vromen and Thomas Wells for helpful discussions and suggestions on previous drafts of the paper. The paper has also benefited from the many insightful comments made by two anonymous referees of the Journal. We also thanks the participants in seminars at the London School of Economics, Université Paris Ouest Nanterre La Défense, Università degli Studi di Milano, Erasmus University Rotterdam, as well as in sessions at the 2009 ASSA meeting, the 2009 STOREP conference and the 2009 HES meeting for their useful comments. Moscati also thanks the Centre for Philosophy of Natural and Social Science at LSE for its hospitality during part of the work on the paper, and is glad to acknowledge the financial support of MIUR PRIN 2007. Any errors are ours.

1. Introduction

Rational Choice Theory (RCT) designates the theory according to which individuals choose among available options to maximize their utility. More extensively, RCT assumes a decision-maker to be rational if her preferences are complete and transitive, and if she chooses her most preferred option. Preferences are not restricted with respect to their orientation, and can be selfish, altruistic, or determined by social norms. If continuous, the preferences of a rational decision-maker can be represented by a utility function, and her choices can be viewed as the result of utility maximization.[1] This is the basic version of RCT that applies to choices not involving risk, studied in individual demand analysis. In the paper we refer mainly to this basic, risk-free version of RCT, although some of our arguments also concern more sophisticated versions of the theory that apply to contexts involving uncertainty, strategic interactions or intertemporal decisions.

Since at least the 1950s, decision theorists have investigated the relationships between rational choice and an apparently very dissimilar behavior, namely, random behavior. One approach employs stochastic models, which extend RCT by adding a random element to rational choice and which can be classified into Fechner Models and Random Preference Models.[2] Fechner Models hypothesize that some noise ε, due to mistakes, inattention or careless evaluation may enter into the individual's decision. As a consequence, although the individual has well-behaved true preferences represented by a utility function U, her choice depends on a perceived utility function $V = U + \varepsilon$, where the noise ε is taken as a zero-mean random variable.[3] In Random Preference Models the individual is characterized by an entire set of well-behaved preference systems, each of which expresses her tastes in a certain state of mind i, and is represented by a utility function U_i. For any decision, one particular utility function among those characterizing the individual is drawn at random, and she chooses the preferred option according to the selected U_i. Both types of stochastic extensions of RCT can accommodate a number of observed violations of the standard, non-stochastic version of RCT. However, each has specific behavioral implications, and hence accounts for distinct violations of RCT.

While stochastic models attempt to integrate rational choice and random behavior, an alternative approach contrasts rational choice with a specific model of random behavior, and uses the latter to discuss the validity of RCT. In this model, a decision-maker chooses at random when he picks up an alternative in his choice set according to some probability distribution on the set, typically a uniform distribution where each alternative in the set has an equal probability of being selected. The motivation for contrasting rational choice with random behavior is that the latter seems to lack any element of rationality and thus to diverge from rational choice more radically than other forms of behavior. Thus, if an economic phenomenon can be explained both as the result of a random and a rational decision, this suggests that it is not rationality itself that matters but other factors unrelated to it which need to be investigated.

Random behavior in the sense just specified was introduced by Alchian (1950), who depicted the economic system as a selection mechanism in which the survival of a firm might be the outcome, not only of rational choices motivated by profit maximization, but also of random behavior accompanied by luck. Adapting Alchian's model to demand analysis, Becker (1962) considered random consumers who pick up a bundle from among those that exhaust their whole income, i.e., a bundle on their budget hyperplane, so that each bundle on the hyperplane has an

equal probability of being chosen. We call this model "Becker random behavior" and focus on it in the present paper.

Becker pointed out that some implications of RCT for demand behavior are, on average, still valid even if consumers choose at random. Inspired by this result, Gode and Sunder (1993) showed that the prices and quantities obtained in double auction markets populated by artificial random traders converge towards the equilibrium price and quantity predicted by the standard supply-and-demand model, thereby suggesting that market efficiency may be largely independent of the rationality and motivation of the agents. More recently, laboratory experiments such as those performed by Cox (1997), Sippel (1997), Mattei (2000), Harbaugh, Krause and Berry (2001), Andreoni and Miller (2002) and Février and Visser (2004), used Becker random behavior as an alternative hypothesis in testing the extent to which individual demand behavior satisfies RCT.

The present paper makes both historical and methodological contributions related to the random-choice model introduced by Becker. Regarding the historical side, despite the importance of Becker's model in demand analysis, to the best of our knowledge there exists no detailed reconstruction of the ways in which it has been used in this area of economics. In section 2 we provide this reconstruction and discuss how the interpretation and use of Becker random behavior have evolved between the original 1962 article and the modern experimental literature on individual demand. In section 3, we also survey recent experiments on individual demand, with a focus on how the predictive success of RCT is contrasted with Becker random choice.

Becker interpreted his result that RCT and random choice have similar implications as a strengthening of RCT, suggesting that the latter provides a compelling explanation of consumer behavior even if consumers choose at random. In making this point, Becker adopted the 'as-if' methodological view popularized in economics by his Chicago mentor, Milton Friedman, in his "The methodology of positive economics" (1953). Becker (1962, p. 4) wrote: "Households can be said to behave not only 'as if' they were rational but also 'as if' they were irrational: the major piece of empirical evidence justifying the first statement can equally well justify the second."

The methodological contribution of the paper consists in an assessment of the as-if defense of RCT suggested by Becker and others. In section 4, we call attention to the fact that the as-if defense only asserts RCT to be a "possible" explanation of demand behavior. This neither rules out the possibility that different models of choice may provide an alternative explanation, nor proves that RCT offers the "best" explanation among the available ones. In this specific sense, we argue, the as-if defense of RCT is a "weak" one. We also claim that supporters of RCT often tend to forget this "weakness" and to contend as if convinced that the as-if argument is *per se* sufficient to endorse RCT. This forgetfulness may explain why they rarely bother to show that RCT accounts for observed demand behavior better than other competing models of choice. In section 5 we argue that, although the as-if defense appeals to RCT's fit with observed behavior, the individual demand experiments surveyed in the paper nevertheless suggest that this fit is poor. In section 6, we further point out that the demand experiments surveyed in the paper were not designed to compare RCT with other possible explanations of demand behavior, but rather posited RCT as the unique null hypothesis under examination and only attempted to reject or validate it. We argue that this is a limitation in their experimental design, and again trace back

this limitation to the widespread tendency to forget the "weakness" of the as-if defense of RCT. Finally, we claim that the reasons for the dominance of the reject-or-validate-RCT approach in the experimental research on individual demand are vanishing.

Our focus on Becker random choice is not only motivated by an historical interest in better understanding its changing role in the last fifty years of demand analysis. More importantly, we are convinced that methodological discussion in economics is more fruitful when anchored to some historical case study. We find that the use of Becker's model as an alternative to RCT in demand analysis constitutes a valuable case study that provides a good, even though not unique standpoint for a methodological assessment of the as-if defense of RCT. Thus, we do not see the methodological contribution of the paper to be independent of the historical one; on the contrary, the former is rooted in the latter.

Some final specifications on the scope and goals of the paper are in order. To begin with, it is important to stress that we claim neither that human consumers generally choose at random, nor that Becker's model explains human demand behavior better than RCT or other choice models. We focus on it because it has historical interest and offers a fruitful methodological standpoint to assess the as-if defense of RCT.

Although the paper focuses on basic RCT that applies to choices not involving risk, it also bears upon the more sophisticated versions of RCT dealing with uncertainty, strategy and time, as well as upon the extensions of RCT that attempt to capture phenomena such as the role of experience and social forces in shaping preferences. In fact, these latter models also construe decision-making as the result of the maximization of some kind of utility function and are interpreted as as-if explanations of human choice behavior. Therefore, our claim that *per se* the as-if defense of RCT is a "weak" one is also relevant for sophisticated and extended RCT.[4]

It may be argued, however, that nobody believes in basic, risk-free RCT anymore, and that we are therefore flogging a dead horse. Basic RCT, however, is still to be found as part of almost any elementary or advanced textbook in microeconomics, and as such seems far from dead. More importantly, our target is not basic RCT but the as-if defense of RCT and the methodological torpor the as-if argument tends to produce among RCT supporters. These latter appear to be alive and well in mainstream microeconomic theory.

2. Random behavior from Becker 1962 to demand experiments

This section investigates how the interpretation and use of Becker random behavior have evolved between Becker's original 1962 article and the modern experimental literature on individual demand.

2.1 Becker's expected random choice

Under the additional assumption of locally non-satiated preferences, RCT implies that a rational consumer chooses a bundle on her budget hyperplane and, among other things, that her demand displays a negative substitution effect.[5] This means that when prices change and the rational consumer is compensated so that at the new prices she can just afford the bundle she chose at the

old prices, her demand for the relatively dearer commodities will decrease.[6] A negative substitution effect is a necessary condition for seeing the consumer's choices as if generated by utility maximization. The condition is not sufficient because substitution effects involve only compensated price changes and a consumer could violate RCT when uncompensated price changes occur.

Becker (1962) imagined a random consumer who chooses a bundle on her budget hyperplane (which with two commodities amounts to a line) according to a uniform distribution, and compared the implications of this random decision rule with those of RCT. As a random consumer has an equal chance of choosing any bundle on the budget line, on average she is expected to pick up the bundle laying at the midpoint of the line and thus to spend half her income on each commodity. In Figure 1, the initial budget line is labeled as AB and e_0 is the midpoint of AB. If commodity x becomes dearer with respect to commodity y, the compensated budget line CD passing through e_0 is steeper than AB. As a result e_1, the midpoint of CD, which is the random consumer's new expected choice, is to the left of e_0. This implies that the expected compensated demand of x has decreased. In other words, the change in relative prices alone shifts consumption opportunities away from relatively more expensive commodities and towards cheaper ones. Therefore, not only RCT but also the random-choice model implies that the substitution effect is negative, at least on average.

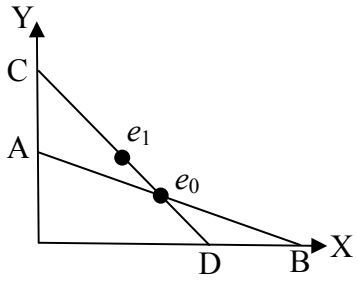

Figure 1: Random choice implies a negative expected substitution effect

As we will see more clearly in a moment, this does not mean that the substitution effect will always be negative. But Becker was chiefly interested in market demand, rather than in individual consumer behavior. RCT predicts that the market demand curve is almost certainly negatively sloped, that is, that the market demand for a commodity and its price move in opposite directions.[7] Becker noted that the random-choice model also implies the near certainty of a negatively sloped market demand curve. In fact, the larger the number of random consumers in the market, the more likely it is that any individual erratic behaviors will cancel out and that the average market demand will approximate the individual expected demand. Hence, a market populated by a large number of random and uncorrelated consumers will almost certainly display a negatively sloped demand curve.

2.2 The limitations of Becker's result

Becker's conclusion that both rational and random consumers display a negative substitution effect draws upon his identification of *random choice* with the *expected outcome of random choice*. However, random choice means that every bundle on the budget line has an equal probability of being chosen, not that the midpoint bundle will always be chosen. In effect, if random choice is reduced to its expected outcome, a random consumer is indistinguishable from a rational consumer with a Cobb-Douglas utility function of the form $U(x, y)=x^{½}y^{½}$, as both pick up the bundle lying at the midpoint of the budget line.

When we consider each and every random choice, by contrast, and rule out the special case of consumers who are totally indifferent between alternatives and thus rational in picking one at random, the convergence between rational and random behavior evaporates. In particular, random behavior no longer implies a negative substitution effect. The compensated budget line CD in Figure 1 shows that there is a probability α equal to the ratio between the length of segments e_0D and CD that the random consumer chooses a bundle to the right of e_0. Therefore, the random-choice model states, in effect, that there is a probability α that the consumer displays a positive substitution effect which violates RCT.

In contrast to Becker, later experimental studies do take into account each and every choice made by random agents, and not only the expected or average outcome of their choices. Becker referred to the expected outcome of random choice because his main interest lay in the negative slope of the aggregate demand curve, rather than in the properties of individual demand, and the expected outcome is sufficient to obtain a negatively sloped market demand curve. The focus of experimental studies, by contrast, is primarily on individual demand and the rationality issues involved, so that random behavior becomes relevant to these topics only if all choices made by random consumers are considered.

Notice, also, that Becker compared the behavioral implications of RCT with those of the random-choice model, but did not test the two theories against the demand behavior of human subjects. When he published his 1962 article, experimental research was still only a niche within economics.[8] Systematic efforts to test RCT in experiments on human demand behavior began only in the 1970s, and most of the research has been performed since the mid-1990s.

2.3 Random behavior in demand experiments

In experiments on individual demand, each human participant is typically asked to choose the preferred bundle among those affordable to her under different budget/price situations, that is, with different incomes and for different commodity prices. The experimenter records subjects' choices and usually checks whether they satisfy the Generalized Axiom of Revealed Preference (GARP).[9] GARP requires that, if the subject directly reveals a preference for bundle e_0 over bundle e_1, that is, if she chooses e_0 when e_1 is available and costs no more than e_0, and if she subsequently directly reveals that she prefers e_1 to e_2, e_2 to e_3, …, and e_{n-1} to e_n, then the subject cannot choose e_n when e_0 is available and costs strictly less than e_n. GARP characterizes RCT in the sense that the choices of a subject can be seen as if generated by the maximization of a locally non-satiated utility function if and only if they satisfy GARP. Therefore, while a negative substitution effect is only a necessary condition of seeing choices as if generated by RCT,

GARP is a necessary and sufficient condition.[10]

Figures 2a-2e below provide a rough geometrical intuition as to which choices satisfy GARP and which violate it in the two-commodity case.[11] In all figures, two budget/price situations are considered: the first identified by budget line AB, and the second by budget line CD. It is assumed that the subject chooses bundle e_0 in situation AB while she picks up e_1 in situation CD. The choices represented in figures 2a, 2b and 2c satisfy GARP, while those in figure 2d violate it. Here, in situation AB the subject directly reveals that she prefers e_0 to e_1, but in situation CD she chooses e_1 although e_0 costs strictly less than e_1. Figure 2e represents the case in which the two budget/price situations coincide but the subject chooses e_0 in AB and e_1 in CD. This pattern of choice can be interpreted as a manifestation of the subject's indifference between e_0 and e_1 and does not violate GARP, since neither of the two bundles costs strictly less than the other.

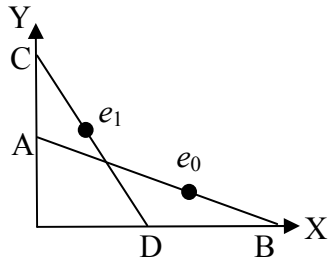

Figure 2a: GARP satisfied

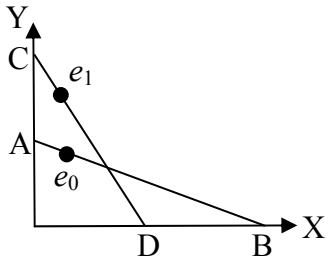

Figure 2b: GARP satisfied

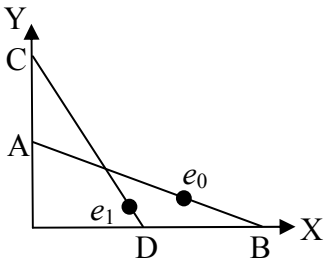

Figure 2c: GARP satisfied

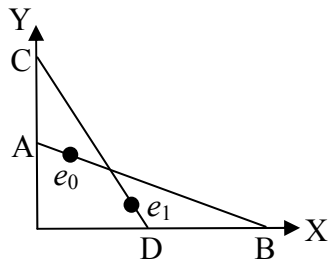

Figure 2d: GARP violated

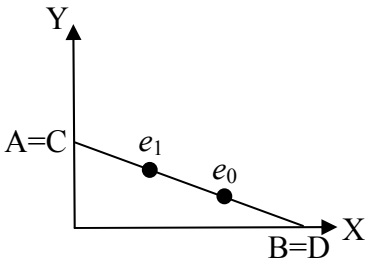

Figure 2e: GARP satisfied

There are, however, two problems with GARP as a test for RCT. First, in order to count a violation of GARP recorded in an experiment as a violation of RCT one needs to assume a form of separability between the demand choices the subjects made within and outside the experiment. Experimental studies typically circumvent the problem by assuming this form of separability more or less implicitly.[12]

Second, even if all choices recorded in the experiment pass the GARP test, this finding may provide little support for RCT. Indeed, human subjects may satisfy GARP simply because in the budget/price situations they face it is extremely difficult or even impossible to violate it. For example, GARP violations become unlikely when the budget hyperplanes intersect near the axes, and impossible when they intersect on the axes or do not intersect at all.

In statistical terms, this is the problem of the power of a test, that is, the probability of a test to reject the so-called null hypothesis (in our case RCT) when the hypothesis is false. To assess the

power of GARP as a test for RCT it is necessary to formulate an alternative hypothesis about the decision rule that could have generated human choices. Becker's random-choice model enters the scene as the alternative hypothesis usually employed in experiments to assess the power of GARP as a test for RCT. The idea is that the lower the probability of GARP violations under random behavior, the lower the power of GARP. In this case even if RCT is false and human subjects choose at random, they would rarely violate GARP.[13]

In order to measure GARP's power yet another obstacle has to be overcome. When consumers choose among more than two commodities and in more than two budget/price situations, calculating the a priori probability of GARP violations under random behavior is extremely difficult. To circumvent this problem, experimenters use Monte Carlo methods to create a population of artificial Beckerian random agents who face the same budget/price situations that the human participants were presented with in the experiment. In each situation, each random agent chooses a bundle on its budget hyperplane according to a uniform distribution. Its choices over the entire set of budget/price situations may or may not violate GARP. The *percentage of random agents* that violate GARP is adopted as a proxy for the a priori probability of GARP violations under random behavior, and hence as a measure of the power of GARP test. If a small proportion of random agents violate GARP, then the fact that human subjects rarely do so provides little support for RCT, since this may be due to the objective difficulty of violating GARP under the budget/price combinations of the experiment rather than due to the rational behavior of participants.[14]

It is important to stress that in checking for GARP violations all choices made by random agents are taken into account, not simply – as in Becker's article – the expected or average outcome of their choices. As observed in Section 2.2, when each and every random choice is considered, it turns out that random agents frequently violate GARP, so that Becker's suggestion of a possible convergence between random and rational individual behavior ceases to be appropriate.

The relative mildness of GARP violations presents a further issue. In all experiments a number of human choices violate GARP, which would imply that experimental data falsify RCT. However, in many cases GARP violations are rare, in the sense that nearly all subjects satisfied GARP, and not severe, in the sense that violators were "close" to passing GARP. In these cases it may be doubted whether it is indeed appropriate to entirely reject RCT. For instance, adding some noise to rational decisions as in stochastic extensions of RCT *à la* Fechner could suffice to rationalize these minor violations of GARP.

Building on the work of Afriat (1967, 1972), Varian (1991) addressed this problem and proposed a weakening of GARP called GARP(*e*), which takes into account the severity of violations through a parameter *e* called the Afriat Efficiency Index. This Index measures the extent to which the budget constraints should be modified in order to accommodate for observed GARP violations. Figure 3 shows that choices e_0 and e_1 in situations AB and CD violate GARP, but, if in situation CD income is reduced so that the modified budget line C'D' passes through e_0, then e_0 and e_1 no longer violate GARP.

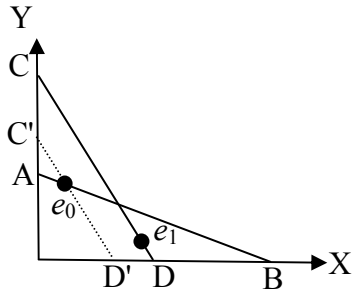

Figure 3: GARP(*e*)

The ratio between income C'D' and income CD expresses the income reduction needed to let GARP violations "disappear", and can be interpreted as an indicator of the money the subject wastes by not choosing rationally. The Afriat Efficiency Index has a maximum value of one (in which case the subject satisfies GARP and no income reduction is needed), decreases as violations become more severe, and has a minimum value of zero. GARP(*e*) with an Afriat Efficiency Index of 0.95, that is, GARP(0.95), can be associated with an acceptable 5% waste of income and is the measure usually considered in the literature. So, for example, if 99% of subjects satisfy GARP(0.95) then they may be seen as "almost rational", in the sense that a minor slackening of the budget constraints would render the near totality of subjects consistent with RCT.

As with GARP, the GARP(*e*) test also present a power problem. If human subjects satisfy GARP(*e*), this may not follow from their almost rational behavior, but rather from the fact that in the budget/price situations they face violations are unlikely. As in the case of GARP, Becker random behavior is used to evaluate the power of GARP(*e*) as test for "almost-RCT".

To summarize, there has been an evolution in the use of the hypothesis of random behavior between Becker's original 1962 article and the modern experimental literature on individual demand. The use of random behavior as an alternative hypothesis to assess the power of GARP as a test for RCT, the reference to the percentage of random violators to measure this power, and the count of each and every choice made by random agents are all major differences between Becker and later experimentalists. Before turning to review, in section 3, the results of six experiments on consumer demand that employed Becker's random behavior and the GARP-GARP(*e*) machinery to test the validity of RCT, we first of all complete our discussion of the changes in the interpretation of Becker's original model by examining some variants upon it proposed in the post-1962 literature.

2.4. Variants of Becker random behavior

The conception of random behavior as the selection of an alternative according to a probability distribution on a choice set allows for some degree of freedom in the identification of the choice set and the specification of the probability distribution. Becker identified the choice set with the set of bundles on the budget hyperplane and employed a uniform probability distribution. The subsequent literature has proposed variations of his original model in which changes are made to the choice set or to the probability distribution on it.

Chant (1963) put forward a version of random behavior in which the decision-maker chooses among goods rather than bundles, in such a way that any unit of the available goods has an equal chance of being selected, independent of its price. Février and Visser (2004) used Chant's version of random behavior in addition to Becker's in experiments on individual demand that aim to assess the power of GARP. Their article, however, exemplifies the main problem with Chant's random agents, namely, that they violate GARP more frequently than do Becker's random agents, so that bringing them into play tends to overrate the power of GARP as a test for RCT.

Even if the inquiry remains confined to random choice among bundles, one may think of the random consumer as picking a bundle from among those she can afford, that is, not only from among those *on* her budget hyperplane, but also from among those *below* it. In effect, restricting the choice set to the budget hyperplane imposes a certain degree of rationality on random agents: if they select a bundle on the hyperplane they are at least avoiding any waste in income.[15] But this restriction seems at odds with one of the main motivations for focusing on random choice in demand analysis, namely, that the latter appears as the opposite of rational choice. However, this restriction has been seldom, if ever, dropped. The reason is that RCT is typically studied and tested under the additional assumption of local-non-satiation, which implies that decision-makers choose a bundle on their budget hyperplane. Requiring random agents to do the same enables a straightforward *ceteris paribus* comparison between RCT and the random-choice model.

Regarding the probability law governing the choices of random agents, Bronars (1987), Cox (1997) and others have proposed various versions of random choice in which the probability law mimics the empirical distribution of observed human choices. We briefly discuss the "bootstrapping" random-behavior model, which has been used in addition to Becker's, in two of the experiments reviewed in section 3, namely, those of Harbaugh, Krause and Berry (2001) and Andreoni and Miller (2002). Imagine an experiment where ten human subjects are confronted with a given budget/price situation. Note that only some bundles among the infinitely many on the budget hyperplane will be chosen by at least one subject, and assume in particular that the subjects' choices concentrate on bundles e_1, e_2, and e_3. For example one subject selects e_1, five subjects pick up e_2, and the remaining four choose e_3. In the bootstrapping random-choice model, the choice set is restricted to $\{e_1, e_2, e_3\}$, and random agents pick up e_1, e_2, and e_3 with a probability of 0.1, 0.5 and 0.4, respectively.

Bootstrapping aims at providing an alternative hypothesis with which to measure the power of GARP more effectively than is possible with Becker's random choice. By conflating the choices of different individuals, however, it creates a sort of representative random agent, whose nature and interpretation are far from clear. Furthermore, Andreoni and Harbaugh (2008) have pointed out that bootstrapping random agents may violate GARP more or less frequently than Becker's random agents, which undermines the original motivation for the bootstrapping technique.

In conclusion, owing to their limitations the variants of Becker random behavior have been but rarely used in the literature, and with all its simplicity Becker's original model has so far dominated experimental research on individual-demand RCT.

3. Review of demand experiments

In this section we briefly review the results of six experiments on consumer demand in which Becker's random behavior was used as an alternative hypothesis against RCT, while in sections 5 and 6 below we discuss their import for our research question. Although we are not aware of other experimental studies of RCT in settings without uncertainty or intertemporal decisions, our review may be not exhaustive. However, the six experiments examined here are sufficient to bring to light the methodological points relevant to the paper.

3.1. Cox (1997) tested GARP violations using data collected by Battalio, Kagel and others (1973) in an experiment conducted with 38 female patients at a mental hospital in Long Island. The patients were part of a token economy established in the hospital, that is, they could earn tokens by performing janitorial tasks and use them to purchase goods sold within the hospital. By varying the token prices of goods, the experimenters created 7 different budget/price situations. The patients could choose from among 16 goods that were grouped into 3 main categories.[16] Table 1 shows the percentage of GARP violations for patients and random agents, and the percentage of GARP(e) violations for e=0.95 for patients only.

Table 1 – Cox 1997

Group	Percentage of subjects violating GARP	Percentage of subjects violating GARP(0.95)
Patients	36.8%	13.1%
Random agents	48.2%	n.a.

A significant fraction of patients (almost 37 percent) violated GARP and hence RCT. The power of the GARP test is not particularly high, as only around 48 percent of random agents violated GARP. The fraction of human violators shrinks to around 13 percent when GARP(0.95) is considered, but the import of this information is not clear since what happens to random violators with GARP(0.95) is unknown.

3.2. Sippel (1997) tested RCT against the demand behavior of 42 students in law or economics at the University of Bonn, Germany. The students were offered 8 food and leisure goods in 10 different budget/price situations.[17] Sippel ran two experiments, involving 12 and 30 subjects, respectively. Table 2 presents the aggregate outcome of the two experiments.

Table 2 – Sippel 1997

Group	Percentage of subjects violating GARP	Percentage of subjects violating GARP(0.95)
Law/Eco students	57.1%	9.5%
Random agents	> 61%	> 12.8%

In Sippel's experiments 24 out of 42 students, or 57 percent, violated GARP. The number of GARP violators among random agents, and hence the power of the GARP test, is above 61

percent in both experiments. When GARP(0.95) is considered, the proportion of random violators, but also the power of GARP(0.95), shrinks noticeably.

3.3. Mattei (2000) studied the demand behavior of 20 microeconomics students (experiment 1), 100 business students (experiment 2), and 320 readers of a consumer affairs magazine (experiment 3). In all three experiments the subjects were faced with 8 goods and 20 different budget/price situations.[18] Table 3 shows the findings.

Table 3 – Mattei 2000

Experiment #	Group	Percentage of subjects violating GARP	Percentage of subjects violating GARP(0.95)
Experiment 1	Micro students	25%	0%
	Random agents	99.4%	43.2%
Experiment 2	Business students	44%	4%
	Random agents	98.9%	43.1%
Experiment 3	Magazine readers	32%	2%
	Random agents	98.9%	42.8%

Mattei also recorded a significant proportion of GARP violators among human subjects, between 25 and 44 percent. The power of the GARP test is excellent, for in each experiment more than 98 percent or random agents violated GARP.

3.4. Harbaugh, Krause and Berry (2001) tested whether children choose rationally. They studied the demand choices of 31 second-grade students aged about 7 years, 42 sixth-grade students aged about 11 and, for comparison, the demand choices of 55 college undergraduates aged about 21. In 11 different budget/price situations children and undergraduates were presented with choice sets including between 3 and 7 bundles. Each bundle contained only 2 commodities, namely potato-chip bags and boxes of fruit juice. Table 4 shows the percentage of subjects violating GARP as well as the Afriat Efficiency Index indicating how much the budget constraints should be relaxed to eliminate GARP violations for the three age groups and for random agents.

Table 4 – Harbaugh, Krause and Berry 2001

Group	Percentage of subjects violating GARP	Afriat Efficiency Index
Second graders, age 7	74%	0.93
Sixth graders, age 11	38%	0.96
Undergraduates, age 21	35%	0.94
Random agents	> 98%	0.648

Harbaugh, Krause and Berry found a significant portion of GARP violators in all three age groups. From age 7 to 11 the number of violators significantly decreases, while from age 11 to

21 it remains more or less at the same level. In no age group were GARP violations particularly severe since they disappear by mildly relaxing the budget constrains.

3.5. In the experiment performed by Andreoni and Miller (2002), 142 students in economics at the University of Wisconsin and Iowa State University had to divide a given number of tokens between themselves and another subject. The tokens were transformed into money, but at different exchange rates. For example, each token a subject kept for himself became $0.10 while each token he passed to another subject became $0.30, or vice versa. Therefore, an individual had in effect to allocate a given token income between two goods, "money for himself" and "money for another subject", such that the relative price of the two money-goods could be larger or smaller than one. By modifying the number of tokens to be divided, and the exchange rates of tokens into money, Andreoni and Miller presented the subjects with 8 different budget/price situations and tested whether their preferences for giving were consistent with RCT. Table 5 shows the findings of the experiment.

Table 5 – Andreoni and Miller 2002

Group	Percentage of subjects violating GARP	Percentage of subjects violating GARP(0.95)
Economics students	9.1%	2.1%
Random agents	78.1%	n.a.

Among the six experiments reviewed, this is the one that records the lowest proportion of human subjects violating GARP. Moreover, the power of the GARP test is satisfactory as more than 78 percent of random agents violated GARP.

3.6. Février and Visser (2004) tested RCT against the demand behavior of 120 individuals from Dijon, France. The subjects were offered 6 different varieties of orange juice in 5 different budget/price situations. Table 6 presents the results of the experiment.

Table 6 – Février and Visser 2004

Group	Percentage of subjects violating GARP	Percentage of subjects violating GARP(0.95)
Dijon consumers	29%	15%
Random agents	22%	5%

In this case GARP violators among random agents turn out to be less numerous than among human subjects. If one considers GARP(0.95), the number of human violators decreases but that of random violators shrinks even more.

4. RCT as possible, rather than best, as-if explanation

Sections 2 and 3 carried out the historical aims of the paper: they traced how the interpretation of Becker's random behavior evolved after his original 1962 article, and reviewed the results of six recent experiments on individual choice, drawing attention to how the predictive success of RCT is contrasted with Becker random behavior. We now turn to the methodological assessment of the as-if defense of RCT.

Many have criticized RCT as being psychologically unrealistic. For instance, Simon and other scholars have contended that real decision-makers lack the cognitive capacities to solve the maximization problem hypothesized by RCT. As a defense against this criticism, supporters of RCT have often adopted the as-if argument of Friedman (1953), which may be presented as follows: it is not the case that actual decision-makers consciously maximize their utility function when choosing; rather, it is the economist who rationalizes the decision-maker's choices as if they were generated by utility maximization. Therefore, the utility function and its maximization are in the mind of the economist rather than the mind of the decision-maker, and so the psychology of the latter is not at issue. Insofar, then, as the decision-maker's observed choices agree with those implicated by RCT, the theory is validated.

But the as-if defense of RCT has been challenged too. Philosophers with sympathies towards scientific realism have disputed the conventionalist or instrumentalist views of scientific theories associated with the as-if argument.[19] Psychologists and behavioral economists have called attention to phenomena such as endowment effects and preference reversals that contradict rather than validate the empirical implications of RCT.[20] We contribute to this debate by pointing out that the as-if defense provides only a "weak" (defined in a specific sense) epistemological support for RCT. Notably, this defense appears "weak" even if one subscribes to a conventionalist or instrumentalist view of scientific theories, so that our point is independent of the controversy concerning the nature of scientific theories.

As mentioned in section 1, Becker interpreted the convergence between RCT and his random-choice model with respect to their implications on demand behavior as a strengthening of RCT. For him, that convergence showed that RCT provides a compelling as-if explanation of consumer behavior even if households choose at random. In section 2.2 we noticed that Becker's convergence results hold only for aggregate demand behavior; but this limitation is irrelevant to the point being made here.

We aim here to call attention to the simple but often overlooked fact that the as-if argument only states that RCT offers *a* possible as-if explanation of (aggregate) demand behavior; it does not rule out the possibility that different models of choice, such as Becker random behavior, may provide an alternative explanation. An as-if-random explanation of demand would go as follows: actual consumers do not choose at random; rather the economist rationalizes their choices as if they were generated by a random process; as far as the consumers' observed choices agree with those generated by the random-choice model, the model is validated. Thus, the as-if argument is in fact a double-edged sword that can be used to support not only RCT, but also Becker's random-choice model, or any other model of choice in accord with observed behavior.

In this sense, the as-if defense alone provides only "weak" epistemological support for RCT: it shows only that RCT offers *a possible* explanation of demand behavior without ruling out other

possible explanations. A "strong" as-if defense of RCT would require showing not only that RCT provides a possible explanation of consumer demand, but also that it offers the *best* available explanation of it. Admittedly, this methodological point could also be made without reference to random behavior, but the as-if-random explanation of consumer demand suggested by Becker allows us to see the issue in a clear way.

The epistemological literature has pointed out that different and often diverging elements may be taken into account in determining which theory emerges as the best explanation. On the one hand, and quite unsurprisingly, the empirical virtues of a theory are important: the better a theory fits with statistical data, experimental findings or even with evidence from introspection, the better it explains. On the other hand, more formal virtues of a theory, such as its simplicity, tractability, or generality, are also relevant in determining its preeminence, even though they may be at variance with the theory's empirical virtues. Without embarking here on a general philosophical discussion of the best-explanation problem,[21] we simply point out that Becker did not address the best-explanation issue and hence did not attempt to show that the empirical or formal virtues of RCT make it the best explanation for observed demand behavior. More importantly, the way Becker used the as-if argument is also the way supporters of RCT tend to employ it: they claim that the theory provides a possible as-if explanation of observed demand behavior but forget the epistemological "weakness" of this argument, and thus rarely bother to show that RCT explains this behavior better than other competing models of choice.

To be clear, we do not argue that Becker random choice explains human behavior better than RCT. Although Becker's model may display some formal virtues such as simplicity and tractability, its empirical qualities are much more problematic. In the first place, introspection suggests that human beings do not usually choose at random. One may imagine that in particular circumstances, e.g. when the available alternatives are very numerous or it is difficult to evaluate and compare them, individuals might actually do so: for instance in the Février-Visser experiment reviewed in section 3.6 subjects had to choose among almost identical commodities (six varieties of orange juice), and this awkward situation may have induced them to pick up alternatives in a random way. But these circumstances appear rare, and from a psychological perspective the scenario of a decision maker who selects an alternative according to a uniform distribution appears even more unrealistic than the image of a decision maker who solves the maximization problem hypothesized by RCT.[22] Secondly, it is easy to imagine experiments that in all probability would provide evidence against Becker's random-choice model. For example, if a decision maker repeatedly faces the same budget/price situation it seems likely that his choices will display much less variance than predicted by Becker's model of random behavior.

Therefore, the contrast between Becker random behavior and RCT appears useful when the goal is to test RCT. As already mentioned, the intuition underlying this contrast is that if an economic phenomenon can be explained both as the result of a random and a rational decision, this suggests that it is not rationality itself that matters but other factors unrelated to it. However when the goal is to find the best explanation for human demand behavior, RCT should be compared with competitors more plausible than Becker random behavior, such as some behavioral or bounded-rationality model of choice. We return to this point in section 6.

In summary, we distinguish between two types of as-if defenses of RCT and, more generally, of a scientific theory. While the "weak" as-if defense states that a theory offers a *possible*

explanation of the phenomenon at hand, the "strong" as-if defense claims that the theory provides the *best* explanation. We argue that the standard as-if defense of RCT is, in this specific sense, a weak one, and that supporters of RCT often tend to forget this "weakness" and indulge in the conviction that the as-if argument is *per se* sufficient to endorse RCT.

To make this point we do not need to dispute the conventionalist or instrumentalist views of scientific theories associated with the as-if argument. Insofar as the conventionalist and the instrumentalist acknowledge that the same phenomenon can be explained or predicted by different theories, they necessarily face the problem of selecting one theory as the best as-if explanation at hand. The weak as-if argument does not address this selection problem, and thus even the conventionalist and the instrumentalist should find the argument wanting.

5. As-if defense of RCT and experimental evidence

In this section we argue that, although the as-if defense appeals to RCT's fit with observed behavior, the six individual demand experiments surveyed in the paper suggest that this fit is poor.

As a preliminary remark, it is important to notice that in the following discussion we assume that the choices recorded in the laboratory are legitimate phenomena against which RCT can be tested, and that we do not dispute the validity of the experimental findings. Indeed, it is always possible to contend that the results obtained in the laboratory rely on the particular experimental design chosen, and that they would disappear under different designs. For example, one may maintain that the six experiments should have been replicated in order to check whether GARP violations were reduced with repetitions; if this were the case, GARP violators would not be irrational but simply requiring of time to get used to choices in the lab. It is also always possible to bring into play the so-called Duhem-Quine problem and contend that experimental results are not significant because some auxiliary assumption necessary to test the theory was not fulfilled.[23] Thus, for example, one may claim that the separability assumption between choices made within and outside the experiment was not met, or that the subjects' preferences changed during the experiment, so that the GARP violations recorded in it should not be interpreted as violations of RCT. Although these and possibly many other criticisms have a point, we think that they indicate the need for further experimental research rather than invalidate the results obtained. Moreover, the six experimental studies reviewed above have been published in top economics journals, and this suggests that the economics profession acknowledges their findings as provisionally sound and relevant for the purpose of testing RCT. In the following discussion, we do the same.[24]

The six experiments show that GARP violators are numerous. Violators range from a minimum of 9.1 percent to a maximum of 74 percent, while in most experiments they are around 30-40 percent. This outcome stands in contrast to RCT, which implies no GARP violations, and thus undermines that part of the as-if defense of RCT relying upon the presumption that the theory is validated by observed choices.

At best, the as-if argument could be used to defend an almost-RCT claiming that "most" individuals choose in a "nearly" rational way, where "nearly" might be interpreted in the sense

of a stochastic extension of RCT *à la* Fechner. In effect, when GARP(0.95) is considered, human violators range from a minimum of 0 percent to a maximum of 15 percent and in most experiments are below 10 percent. However with GARP(0.95) the proportion of random violators also shrinks significantly, so that the power of GARP(0.95) as a test for almost-RCT is wanting.

A second finding of the experiments is that the number of GARP violators, and thus the consistency of RCT with observed choices, varies strongly from experiment to experiment. In some experiments, such as Andreoni-Miller's, RCT fits observed choices quite well, whereas in others, such as Sippel's, the theory is conspicuously at odds with experimental evidence.[25]

The variance of the fit of RCT may depend upon various factors, such as the amount of time available for the subjects to decide, the finite or nearly infinite number of bundles in their choice sets, or the order in which the different choice sets are presented to them. Yet no general indication can be drawn from the experiments. The only tendency that seems traceable is that the more goods and budget/price situations the subjects face, the more they violate GARP. There are exceptions, though: in Mattei's experiment the subjects were confronted with 8 goods in 20 budget/price situations, while in Sippel's they faced 8 goods in 10 situations; yet Mattei recorded fewer violators than did Sippel. In short, there is no satisfactory meta-theory that states in which situations RCT works and in which it doesn't.

This conclusion is relevant for the best-explanation issue. An often mentioned advantage of RCT over behavioral or bounded-rationality models of choice is that the former is a single theory with an ambition to achieve general explanatory power, while the latter constitute a whole range of different decision rules that are tailor-made to cater for particular contexts but are not always transferable to other sets of circumstances. If the formal virtues of theories play a role in the determination of which emerges as the best explanation, then RCT's generality, which contrasts with the particularity of the alternative models, seems to give it an advantage over these alternatives. However, the experimental finding that RCT satisfactorily fits certain patterns of choice but not others suggests that it has at best local rather than general explanatory power. This renders RCT more similar to behavioral and bounded-rationality models of choice than is usually assumed, and undermines the argument that RCT's generality compensates for its empirical shortcomings and justifies its selection as the best explanation at hand.

6. As-if defense of RCT and limitations in experimental design

In Cox's study, in Sippel's experiments, and in the test performed by Février and Visser, the proportion of GARP violators among human subjects is significantly different from the proportion predicted by RCT, i.e. zero, and is instead quite similar to the proportion of random GARP violators. One may be tempted to infer that, at least in certain situations, Becker's random-choice model fits human behavior better than does RCT. This inference is, however, incorrect. The experimental evidence collected in those three experiments only shows that human subjects and random agents violate GARP (hence RCT) in a similar proportion, not that the two groups behave in a similar way. Indeed, the choice patterns of human and random agents might be highly diverse and the two groups may violate GARP in very different ways. If this is the case, the random-choice model would fit human demand behavior as poorly as RCT.

From a statistical viewpoint this may be seen by noting that the GARP test is constructed to check the null hypothesis that RCT fits human demand behavior, not the alternative hypothesis that Becker's random choice fits it. To investigate this latter issue would require an explicit statistical test in which the null hypothesis is that the recorded choices of each human subject come from a uniform distribution on the subject's choice sets, whereas the alternative hypothesis is that they do not.

As already mentioned, one may imagine that, at least in certain circumstances, subjects choose at random. But the six experimental studies did not investigate this possibility. Becker random behavior remained an implausible benchmark used to measure the power of the GARP test for RCT, and never rose to the role of credible alternative to RCT.[26]

More generally, the experimental studies were not designed to compare RCT's goodness-of-fit with human demand behavior, on the one hand, with the goodness-of-fit of some other competing model of choice, on the other, may this latter be the implausible Becker random behavior or some alternative – and possibly more credible – model. This appears as a methodological limitation of the six experiments.

We suggest that this limitation is related to the conviction that the as-if argument alone suffices to endorse RCT: in so far as the only question at stake is whether RCT provides a possible explanation of demand behavior, the experiments will aim at falsifying or validating RCT and will posit RCT as the unique null hypothesis under examination. It is only if one wants to make a strong case for RCT and to show that it offers the best explanation of consumer behavior that the goodness-of-fit of different competing choice models will be investigated.

The search for the best explanation is more common in the study of individual choice under uncertainty or in intertemporal settings, that is, in contexts beyond the domain of the individual-demand version of RCT considered in this paper, and in other areas of experimental economics. For instance, already in the 1970s Fiorina and Plott (1978) used experiments to investigate which among different competing models of committee decisions performed best. Plott (1986, p. S302) expressed the view that, for those who study experimental markets, "the rejection of a theory […] is not an especially challenging research objective"; rather, for them "the question becomes which of several competing models is the most accurate, fully realizing that the best model might still be 'poor'". Again, in a study of behavior under uncertainty, Bateman, Kahneman and others (2005) experimentally tested competing theories of loss aversion, while Halevy (2007) used an experiment to compare alternative theories of ambiguity aversion. In the analysis of intertemporal decisions, Manzini and Mariotti (2010a) tested RCT against a bounded-rationality model they called "Categorize Then Choose", and concluded that it fits experimental data better than RCT.[27]

The divergence between the reject-or-validate-RCT approach typical of experimental research in risk-free individual demand and the search-for-the-best-explanation approach of other areas of experimental economics is to be attributed to various factors. While in demand analysis the primacy of RCT has always been so firm that it might have appeared as if there were no alternatives to test, in the analysis of committee decisions, loss aversion, ambiguity aversion or intertemporal decisions a bulk of competing models have been put forward. This proliferation of models has probably stimulated experimental research aimed at comparing them. Moreover, for a long time in the study of individual choice without uncertainty the main alternative to RCT has

been represented by bounded-rationality models *à la* Simon. Although these models were introduced on the basis of plausible psychological insights, their behavioral implications too often remained unspecified. This circumstance made their comparison with RCT difficult and thus contributed to maintaining an exclusive attention on RCT. Finally, as already argued, the focus of demand experiments on the rejection or validation of RCT can be also traced back to the (misplaced) conviction that demonstrating that RCT provides a possible as-if explanation of consumer behavior already constitutes a robust defense of it.

These factors, however, seem to be vanishing. Newer models of bounded-rationality, such as those put forward by Masatlioglu and Ok (2005) and by Manzini and Mariotti (2007, 2009, 2010b), have precise behavioral implications that facilitate comparison with RCT. In the last two decades behavioral economics has called into question the supremacy of RCT, and it has done so more effectively than any of the previous criticisms, some of which date back to the late-nineteenth century. Finally, in this paper we hope to have exposed the epistemological weakness of the standard as-if defense of RCT. Thus we anticipate that future experimental research in demand analysis will abandon the reject-or-validate-RCT approach, and adopt the epistemologically sounder search-for-the-best-explanation method.

7. Summary and conclusions

In this paper we have offered both a historical and a methodological contribution, the two being linked by the role played in each by Becker random behavior as an alternative hypothesis to RCT.

At the historical level, we have investigated how the interpretation and use of Becker random behavior evolved between the original 1962 article and the modern experimental literature in individual demand, and we have surveyed six recent experiments in which Becker's random behavior has been used in testing RCT. Among other things, we have pointed out that while Becker was interested in RCT's implications for aggregate demand and thus focused on the expected or average outcome of random choice, later experimentalists investigated RCT's implications for individual demand and took into account each and every choice made by random agents. With this latter approach, it is apparent that random agents frequently violate RCT and so Becker's convergence results between rational and random behavior fade away. In effect, the experimentalists used random behavior only as an alternative hypothesis to assess the power of GARP as a test for RCT, referred to the percentage of random violators to measure this power, and never investigated the possibility that human consumers choose at random. Finally, we have examined some variations of Becker random behavior as proposed in the post-1962 literature and concluded that these variations lack the features that render Becker's original model an implausible but useful benchmark for testing RCT.

The historical assessment of Becker's model of random choices provided the appropriate standpoint for a methodological evaluation of the as-if defense of RCT. First, the as-if-random explanation of demand suggested by Becker shows that the as-if defense demonstrates only that RCT offers a possible explanation of demand behavior, not that RCT provides the best available explanation. In this definite sense, the as-if defense of RCT is a "weak" one. Supporters of RCT often tend to forget this "weakness" and thus rarely bother to attempt to show that RCT accounts

for observed demand behavior better than other competing models of choice. Second, our survey of six individual demand experiments suggests that RCT fits human demand behavior poorly. This outcome undermines that part of the as-if defense of RCT relying upon the presumption that the theory is validated by observed choices. Moreover, the fact that RCT's fit with consumer behavior varies strongly from experiment to experiment weakens RCT's pretense of generality and thus also its claim to provide the best explanation of human demand behavior. Third, we have pointed out that the six demand experiments were not designed to compare RCT with other possible explanations of demand behavior, but were only attempts to falsify or validate RCT. This is a limitation in their design which seems related to the common conviction that the as-if argument *per se* provides sufficient reason to adopt RCT. Finally, we have argued that the causes for the dominance of the reject-or-validate-RCT approach in the experimental research on individual demand are vanishing, and we have anticipated that future research in this field will adopt the epistemologically sounder search-for-the-best-explanation method.

Notes

[1] Preferences are continuous if they do not present sudden reversals. More formally, an individual has continuous preferences when her preferring each element in a sequence of options $\{x^n\}$ to the corresponding element in the sequence $\{y^n\}$, implies that she also prefers the limiting option of the first sequence to the limiting option of the second sequence. On continuity and the other properties of preferences, see Mas-Colell, Whinston and Green (1995, chapters 1 and 3).

[2] Stochastic models were introduced by Georgescu-Roegen (1950), Quandt (1956), Luce (1958, 1959) and Marschak (1959). For a discussion of more recent contributions to the research program on stochastic models, see Loomes (2005) and Wilcox (2008). The classification of stochastic models used here follows Loomes and Sugden (1995, 1998).

[3] In applied demand analysis, McFadden (1974) and others have introduced stochastic models of the Fechner type, usually labelled Random Utility Models. In these models the random component ε is also meant to accommodate measurement errors on the part of the observer.

[4] For a presentation of RCT in contexts involving uncertainty, strategy and time see Mas-Colell, Whinston and Green (1995, chapters 6, 8-9, and 20). An example of extended RCT that incorporates experiences and social forces into the theory is provided by Becker (1996).

[5] A consumer has locally non-satiated preferences when for any consumption bundle x there exists another bundle y arbitrarily close to x which is strictly preferred to x by the consumer.

[6] This kind of compensation is called Slutsky's compensation. With Hicksian compensation, by contrast, the consumer is compensated so that her utility level is kept constant when prices change. Both kinds of compensations induce negative substitution effects, but Slutsky's are those used in empirical studies because they can be determined even without knowing the consumer's utility function.

[7] This is the so-called "law of market demand." Exceptions to the law are represented by Giffen goods, which are, however, rare for individual demand and extremely implausible for market demand. For an analysis of Giffen goods in individual and market demand see Battalio, Kagel and Kogut (1991).

[8] For a history of the early experiments on demand behavior, see Moscati (2007).

[9] An earlier strand of experimental research employed pigeons and rats as experimental subjects, and checked whether their demand displayed a negative substitution effect. This research is summarized in Kagel, Battalio and Green (1995).

[10] GARP was introduced by Varian (1982) and is a modification of the Weak Axiom of Revealed Preference (WARP) proposed by Samuelson (1938), and the Strong Axiom of Revealed Preference (SARP) proposed by Houthakker (1950). WARP allows for cyclical choices, which are excluded by RCT, and excludes indifference curves with straight segments, which are compatible with RCT. SARP rules out cyclical choices but still excludes straight indifference curves. GARP rules out cycles and allows for straight indifference curves, thereby providing a complete behavioral characterization of RCT.

[11] The intuition is rough since cyclical choices, which are ruled out by GARP, may materialize only when at least three commodities and three budget/price situations are involved.

[12] On the separability assumption, see Varian (1983, 1988).

[13] For a detailed discussion of the power of the GARP test, see Bronars (1987) and Andreoni and Harbaugh (2008).

[14] In addition to the percentage of random agents that violate GARP, one could also employ the percentage of GARP violations as a measure of GARP's power; indeed such a measure is computed by some experimenters. The main problem with it is that there are different ways to count GARP violations. For instance, choices like those in Figure 2d count as one violation in some experiments, and are regarded as two violations in others. We focus on the percentage of random agents that violate GARP because this measure is univocally determined, because it was calculated in all experiments reviewed in Section 3, and because no significant new insight is gained by combining it with the percentage of GARP violations.

[15] For a critical discussion of the implications of this restriction, see Tubaro (2009).

[16] The goods were: cigarettes, coffee, two types of candy, cookies, soda, milk, meal deal with a cigarette (category one); private dormitory room, private locker, grounds pass to leave the ward for a fixed period of time (category two); repeated use of the ground pass, clothes, weekly dance, breakfast, different rights such as right to use cash for packages from home (category three).

[17] The goods were: Coca-Cola, orange juice, coffee, licorice, snacks, music video clips, computer games, magazines.

[18] In experiment 1 the goods were: milk chocolate, salted peanuts, biscuits, text markers, ballpoint pens, plastic folders, writing pads, post-it notes. In experiment 2, milk chocolate, biscuits, orange juice, iced tea, writing pads, plastic folders, diskettes, post-it notes. In experiment 3, milk chocolate, biscuits, orange juice, iced tea, post-it notes, audiocassettes, ballpoint pens, batteries.

[19] Scientific realism is the view that scientific theories describe the world as it really is. Conventionalism contends that scientific theories rationalize experience in a simple, systematic and possibly conventional way, rather than describing the world as it truly is. Instrumentalism views scientific theories as mere instruments for prediction. For a discussion of realism, conventionalism and instrumentalism, see Bird (1998, chapter 4), Chalmers (1999, chapter 15), and Psillos (1999). For a realist critique of Friedman's as-if argument, see Blaug (1980), Caldwell (1980), Musgrave (1981) and Hausman (1992, chapter 9).

[20] See, for example, Tversky and Thaler (1990), Kahneman, Knetsch and Thaler (1991, 2008), and Seidl (2002).

[21] For a discussion, see among others Thagard (1978) and Lipton (2004).

[22] In game theory, the psychological realism of random behavior has been discussed in relation to mixed strategies. Under the interpretation proposed by Harsanyi (1973), players do not randomly choose between different pure strategies but choose rather a single pure strategy in a game of incomplete information associated with the original game of complete information. For a discussion, see Osborne and Rubinstein (2001, chapter 3).

[23] On the Duhem-Quine problem, see Hands (2001, chapter 3).

[24] Using the terminology introduced by Cubitt (2005), we are assuming that the choices recorded in the experiments belong to the 'Testing-domain' of RCT. For a discussion of the Testing-domain of individual choice theory under conditions of uncertainty, see Bardsley, Cubitt and others (2010, chapter 2).

[25] Satz and Ferejohn (1994) discuss the variance of RCT's explanatory power and its dependency on the particular environment that the theory is applied to. In particular, they suggest that rational-choice explanations are most plausible "under conditions of scarcity, where human choice is severely constrained," while "in environments without strong constraints, agents will not generally behave as the theory predicts" (p. 81). Their conclusions, however, do little to shed light upon RCT's violations and the variance of RCT's explanatory power across the six experiments, in which subjects faced almost the same choice situation and the Satz-Ferejohn scarcity condition for rational choice was satisfied because the subjects' choices were always budget-constrained.

[26] Actually, Cox (1997, p. 1076) suggested a rough test to compare RCT and the random-choice model, and argued that the individual's choices in his experiments were more consistent with RCT than with random choice. However, not even Cox has examined the issue in detail.

[27] In "Categorize Then Choose" (CTC) the agent decides in two stages. First she categorizes the alternatives in broad classes and focuses on one class; then she chooses an alternative from that class. For example, a CTC agent categorizes restaurants by type of cuisine and focuses on, say, Mexican restaurants; then she chooses the preferred Mexican restaurant. CTC agents may violate RCT.

References

Afriat, S. (1967), 'The Construction of a Utility Function From Expenditure Data', *International Economic Review*, 8, 67-77.

Afriat, S. (1972), 'Efficiency Estimates of Productions Functions', *International Economic Review*, 13, 568-598.

Alchian, A.A. (1950), 'Uncertainty, Evolution, and Economic Theory', *Journal of Political Economy*, 58, 211-221.

Andreoni, J., and Harbaugh, W. (2008), 'Power Indices for Revealed Preference Test', mimeo, University of California, San Diego.

Andreoni, J., and Miller J. (2002), 'Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism', *Econometrica*, 70, 737-753.

Bardsley, N., Cubitt, R., and others (2010), *Experimental Economics. Rethinking the Rules*. Princeton and Oxford: Princeton University Press.

Bateman, I., Kahneman, D., and others (2005), 'Testing Competing Models of Loss Aversion: An Adversarial Collaboration', *Journal of Public Economics*, 89, 1561-1580.

Battalio R.C., Kagel, J.H., and Kogut, C.A. (1991), 'Experimental Confirmation of the Existence of a Giffen Good', *American Economic Review*, 81, 961-970

Battalio, R.C., Kagel, J.H., and others (1973), 'A Test of Consumer Demand Theory Using Observations of Individual Consumer Purchases', *Western Economic Journal*, 11, 411-428.

Becker, G.S. (1962), 'Irrational Behavior and Economic Theory', *Journal of Political Economy*, 70, 1-13.

Becker, G.S. (1996), *Accounting for Tastes*, Cambridge, Mass.: Harvard University Press.

Bird, A. (1998), *Philosophy of Science*. Abingdon, U.K.: Routledge.

Blaug, M. (1980), *The Methodology of Economics*. Cambridge, U.K.: Cambridge University Press.

Bronars, S.G. (1987), 'The Power of Nonparametric Tests of Preference Maximization', *Econometrica*, 55, 693-698.

Caldwell, B.J. (1980), 'A Critique of Friedman's Methodological Instrumentalism', *Southern Economic Journal*, 47, 366-374.

Chalmers, A.F. (1993), *What Is This Thing Called Science?*, Maidenhead and New York: Open University Press.

Chant, J.F. (1963), 'Irrational Behavior and Economic Theory: A Comment', *Journal of Political Economy*, 71, 505-510.

Cox, J.C. (1997), 'On Testing the Utility Hypothesis', *Economic Journal*, 107, 1054-1078.

Cubitt, R. (2005), 'Experiments and the Domain of Economic Theory', *Journal of Economic Methodology*, 12, 197-210.

Février, P., and Visser, M. (2004), 'A Study of Consumer Behavior Using Laboratory Data', *Experimental Economics*, 7, 93-114.

Fiorina, M.P., and Plott, C.R. (1978), 'Committee Decisions under Majority Rule: An Experimental Study', *American Political Science Review*, 72, 575-598.

Friedman, M. (1953), 'The Methodology of Positive Economics', in M. Friedman, *Essays in Positive Economics*, Chicago: University of Chicago Press, pp. 3-43.

Georgescu-Roegen, N. (1950), 'The Theory of Choice and the Constancy of Economic Laws', *Quarterly Journal of Economics*, 64, 125-38.

Gode, D.K., and Sunder, S. (1993), 'Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality', *Journal of Political Economy*, 101, 119-137.

Halevy, Y. (2007), 'Ellsberg Revisited: An Experimental Study", *Econometrica*, 75, 503-536.

Hands, D.W. (2001), *Reflection without Rules*. Cambridge: Cambridge University Press.

Harbaugh, W., Krause, K., and Berry, T. (2001), 'GARP for Kids: On the Development of Rational Choice Behavior' *American Economics Review*, 91, 1539-1545.

Harsanyi, J.C. (1973), 'Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points', *International Journal of Game Theory*,,2, 1-23.

Hausman, D. (1992), *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.

Houthakker, H.S. (1950), 'Revealed Preference and Utility Function', *Economica* [NS], 17, 159-174.

Kagel, J.H., Battalio, R.C., and Green, L. (1995), *Economic Choice Theory: An Experimental Analysis of Animal Behavior*, New York: Cambridge University Press.
Kahneman, D., Knetsch, J.L., and Thaler R.H. (1991), 'Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias', *Journal of Economic Perspectives*, 5, 193-206.
Kahneman, D., Knetsch, J.L., and Thaler R.H. (2008), 'The Endowment Effect: Evidence of Losses Valued more than Gains', in C.R. Plott and V.L. Smith (eds), *Handbook of Experimental Economics Results*, Volume 1, North-Holland: Amsterdam, pp. 939-948.
Lipton P. (2004), *Inference to the Best Explanation*, London: Routledge.
Loomes, G. (2005), 'Modelling the Stochastic Component of Behaviour in Experiments: Some Issues for the Interpretation of Data', *Experimental Economics*, 8, 301-323.
Loomes, G., and Sugden, R. (1995), 'Incorporating a Stochastic Element into Decision Theories', *European Economic Review*, 39, 641-648.
Loomes, G., and Sugden, R. (1998), 'Testing Different Stochastic Specifications of Risky Choice.', *Economica*, 65, 581-598.
Luce, R.D. (1958), 'A Probabilistic Theory of Utility', *Econometrica*, 26, 193-224.
Luce, R.D. (1959), *Individual Choice Behavior*, New York: Wiley.
Manzini, P., and Mariotti, M. (2007), 'Sequentially Rationalizable Choice', *American Economic Review*, 97, 1824-1839.
Manzini, P., and Mariotti, M. (2009), 'Consumer Choice and Revealed Bounded Rationality', *Economic Theory*, 41, 379-392.
Manzini, P., and Mariotti, M. (2010a), 'Revealed Preferences and Boundedly Rational Choice Procedures: an Experiment', mimeo, University of St. Andrews.
Manzini, P., and Mariotti, M. (2010b), 'Categorize Then Choose: Boundedly Rational Choice and Welfare', mimeo, University of St. Andrews.
Marschak, J. (1959), 'Binary Choice Constraints and Random Utility Indicators', in K.J. Arrow, S. Karlin and P. Suppes (eds), *Mathematical Methods in the Social Sciences*, Stanford: Stanford University Press, pp. 312-329.
Masatlioglu, Y., and Ok, E.A. (2005), 'Rational Choice with Status Quo Bias', *Journal of Economic Theory*, 121, 1-29.
Mas-Colell, A., Whinston M.D., and Green, J.R (1995), *Microeconomic Theory*, New York: Oxford University Press.
Mattei, A. (2000), 'Full-Scale Real Tests of Consumer Behavior Using Experimental Data', *Journal of Economic Behavior and Organization*, 43, 487-497.
McFadden, D. (1974), 'Conditional Logit Analysis of Qualitative Choice Behavior', in P. Zarembka (ed.), *Frontiers of Econometrics*, New York: Academic Press, pp. 105-142.
Moscati, I. (2007), 'Early Experiments in Consumer Demand Theory: 1930-1970', *History of Political Economy*, 39, 359-401.
Musgrave, A. (1981), '"Unreal Assumptions" in Economic Theory: The F-Twist Untwisted', *Kyklos*, 34, 377-387.
Osborne, M.J., and A. Rubinstein (2001), *A Course in Game Theory*, Cambridge, Mass.: The MIT Press.
Plott, C.R. (1986), 'Rational Choice in Experimental Markets', *Journal of Business*, 59, S301-S327.
Psillos, S. (1999), *Scientific Realism: How Science Tracks Truth*, London: Routledge.
Quandt, R.E. (1956), 'A Probabilistic Theory of Consumer Behavior', *Quarterly Journal of Economics*, 70, 507-36.
Samuelson, P. (1938), 'A Note on the Pure Theory of Consumer's Behaviour', *Economica* [NS], 5, 61-71.
Satz, D., and Ferejohn, J. (1994), 'Rational Choice and Social Theory', *Journal of Philosophy*, 91, 71-87.
Seidl, C. (2002), 'Preference Reversal', *Journal of Economic Surveys,* 16, 621-655.
Sippel, R. (1997), 'An Experiment on the Pure Theory of Consumer's Behavior', *Economic Journal*, 107, 1431-1444.
Thagard, P.R. (1978), 'The Best Explanation: Criteria for Theory Choice', *Journal of Philosophy*, 75, 76-92.
Tubaro, P. (2009), 'Is Individual Rationality Essential to Market Price Formation? The Contribution of Zero-Intelligence Agent Trading Models', *Journal of Economic Methodology,* 16, 1-19.

Tversky, A., and Thaler, R.H. (1990), 'Anomalies: Preference Reversals', *Journal of Economic Perspectives*, 4, 201-211.
Varian, H.R. (1982), 'The Nonparametric Approach to Demand Analysis', *Econometrica*, 50, 945-73.
Varian, H.R. (1983), 'Non-Parametric Tests of Consumer Behaviour', *Review of Economic Studies*, 50, 99-110.
Varian, H.R. (1988), 'Revealed Preference with a Subset of Goods', *Journal of Economic Theory,* 46, 179-185.
Varian, H.R. (1991), 'Goodness of Fit for Revealed Preference Tests', mimeo, University of Michigan.
Wilcox, N.T. (2008), 'Stochastic Models for Binary Discrete Choice under Risk: A Critical Primer and Econometric Comparison', *Research in Experimental Economics*, 12, 197-292.